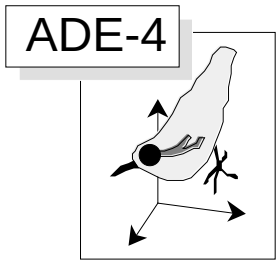


Repères dans la documentation

Résumé

La fiche donne le plan de la documentation thématique. Elle sert à savoir rapidement ce qu'on peut faire avec le logiciel ADE-4.

Plan

1 — Introduction..... 2

2 — Analyses élémentaires..... 5

3 — Variables instrumentales 7

4 — Couplages et contraintes..... 9

5 — Méthodes à K tableaux..... 11

6 — Distances et partition 14

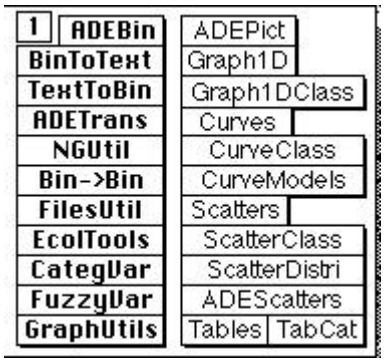
7 — Cartographie 14

Références 16

D. Chessel & J. Thioulouse

1 — Introduction

Chacune des 7 parties du logiciel est symbolisé par une partie de la carte d'entrée. La documentation est de même divisée en sept parties comprenant d'une part la documentation des fonctions des modules, d'autre part des exemples d'utilisation orientés par des données numériques du champ écologique.

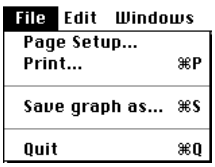


La documentation thématique est réduite aux informations générales d'initiation au logiciel. La documentation des modules est la plus conséquente et comprend les fiches de 16 des 23 modules affichés. On y trouve l'essentiel des options de manipulation de fichiers, des méthodes élémentaires de manipulation de l'information écologique et l'ensemble des modules graphiques de base. Il est conseillé de prendre contact par des exemples avec la **partie de l'interface utilisateur commune à tous les modules graphiques**.

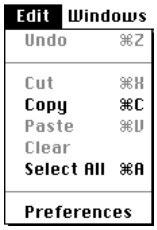
Tous les modules graphiques possèdent trois fenêtres de dialogue non modales : la fenêtre de dialogue principale (File selection), qui est la même que celle des modules de calcul, la fenêtre Min/Max, qui permet de modifier divers paramètres purement graphiques, et la fenêtre de sélection des lignes et des colonnes, qui permet de constituer les collections de graphiques par sélection de colonnes et de groupes de lignes. Les modules graphiques possèdent de plus une fenêtre graphique, dans laquelle les dessins sont tracés. Ils ne possèdent par contre pas de fenêtre de listing d'exécution.

La partie inférieure de la fenêtre de dialogue principale possède une série de boutons permettant d'effectuer trois opérations sur le graphique courant (c'est à dire celui affiché dans la fenêtre graphique): recopie dans le presse-papier, enregistrement dans un fichier et impression.

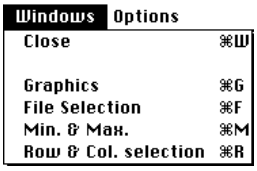
L'impression et l'enregistrement dans un fichier peuvent aussi être réalisés par l'intermédiaire du menu File, qui permet de plus de sélectionner le format d'impression (Page Setup...):



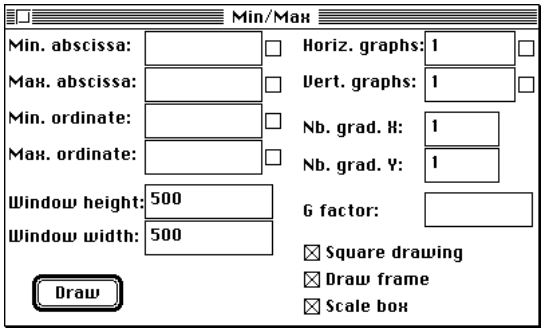
La recopie du graphique courant dans le presse papier peut elle être réalisée par l'intermédiaire du menu Edit, qui donne également accés à la commande Preferences, permettant de choisir le dossier de travail:



Le menu Windows permet quant à lui de sélectionner une des fenêtres:



La fenêtre Min/Max permet de modifier les paramètres suivants :

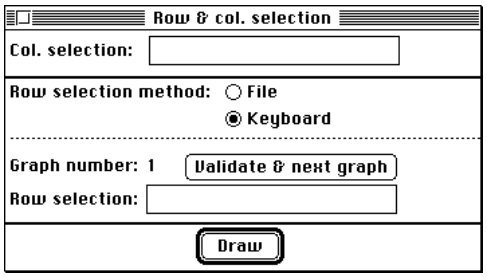


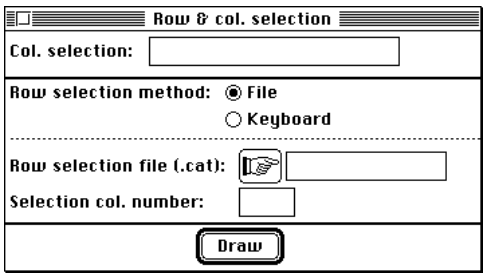
- Minimum et maximum des abscisses et des ordonnées
- Hauteur et largeur de la fenêtre graphique (défaut=500x500)
- Nombre de graphiques horizontaux et verticaux (pour les collections)
- Nombre de graduations sur les axes X et Y (défaut=1 et 1)
- Facteur G (facteur de proportionalité pour les cercles et les carrés)
- Restriction de la zone de tracé à une région carrée dans la fenêtre graphique (défaut=oui)
- Tracé d'un cadre autour de chaque graphique (défaut=oui)
- Tracé d'un cartouche d'échelles (défaut=oui)

Des cases à cocher permettent de plus d'interdire le recalcul automatique des paramètres suivants :

- Minimum des abscisses
- Maximum des abscisses
- Minimum des ordonnées
- Maximum des ordonnées
- Nombre de graphiques horizontaux
- Nombre de graphiques verticaux

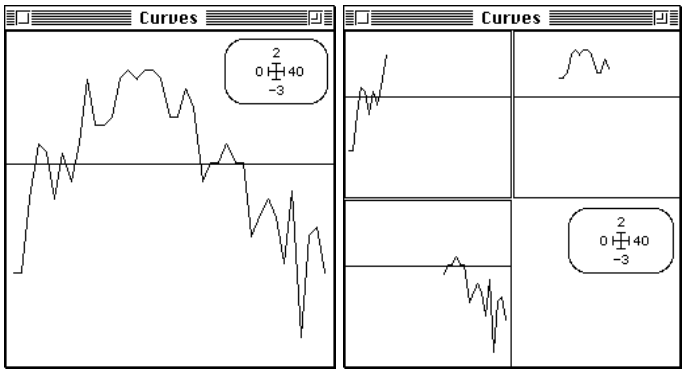
La fenêtre de sélection des lignes et des colonnes possède deux états selon que la sélection des lignes doit être faite grâce à un fichier de sélection ou au clavier :





Pour la sélection des colonnes, il suffit de taper les numéros des colonnes choisies, séparés par des points-virgule (1;3;5). Pour spécifier une plage de valeurs, on peut ne taper que le premier et le dernier élément de la plage, séparés par un caractère non numérique (1-5). Les deux types d'entrées peuvent être mélangées (1;3;5-8).

Dans le cas de la sélection des lignes au clavier (figure ci-dessus en haut), il suffit de taper les numéros des lignes choisies avec la même syntaxe que pour les colonnes (cf. paragraphe précédent), puis de cliquer sur le bouton "Validate & next graph", et de répéter cette opération pour tous les graphiques désirés. Dans l'exemple ci-dessous le graphique de gauche est constitué de 39 points, et il est décomposé en trois graphiques (à droite) en sélectionnant les lignes 1 à 10, 11 à 23, et 24 à 39: taper 1-10 dans le champ "Row selection", puis cliquer sur le bouton "Validate & next graph", taper 11-23, cliquer à nouveau, taper 24-39, et cliquer une troisième fois. A chaque clic, le numéro du graphique est incrémenté automatiquement de un (Graph number:).



Pour la sélection des lignes par l'intermédiaire d'un fichier de sélection, il faut disposer d'un fichier contenant une variable qualitative dont les modalités correspondent aux numéros des graphiques de la collection à réaliser. Pour effectuer le même graphique que ci-dessus, le contenu de ce fichier doit donc être le suivant (10 fois 1, 13 fois 2 et 16 fois 3, sur une colonne) :

1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3

Ce fichier qualitatif doit être traité auparavant avec l'option CategVar : Read Categ File. C'est le fichier dont le nom est muni de l'extension ".cat" qui doit être indiqué dans la fenêtre de dialogue. Il peut posséder plusieurs colonnes, et la colonne à utiliser doit être indiquée (par défaut c'est la colonne numéro un qui est utilisée).

2 — Analyses élémentaires

Cette partie est dite des analyses de premier niveau (un tableau). La documentation des modules est détaillée et contient des indications tant sur la mise en œuvre que sur la signification de nombreuses variantes de l'analyse à un tableau. La documentation thématique porte sur le modèle général du "schéma de dualité" appelée aussi analyse générale d'un triplet statistique et ses variantes : Analyses en composantes principales, Analyse des correspondances simples, multiples, floues, non symétriques, ... les analyses inter et intra-classes, les mélanges de type de variables et les outils graphiques de dépouillement qui sont souvent originaux.

2.1Introduction to Linear Data Analysis

This volume describes the main features of the mathematical model associated with the linear multivariate methods incorporated into ADE-4 software. The main characteristics of statistical computation for one-table analysis are incorporated. ADE-4 is a data analysis package designed for descriptive statistics, which is useful for large data sets to investigate the structure or the organisation of data. Several illustrations of that volume are part of papers from Chessel & Thioulouse (1990, Auto-modélisation en analyse des données. In : Modélisation confluent des sciences. Brissaud, M., Forsé, M. & Zighed, A., Eds., Editions du CNRS. Paris. 71-86), Thioulouse et al. (1991, Graphical techniques for multidimensional data analysis. In : *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers, J. & Karcher, W., Eds., Kluwer Academic Publishers. 153-205), Dolédec & Chessel (1991, Recent developments in linear ordination methods for environmental sciences. Advances in Ecology, India : 1, 133-155), and Chevenet et al. (1994, A fuzzy coding approach for the analysis of long-term ecological data. Freshwater Biology : 31, 295-309).

S. Dolédec & D. Chessel. 32 pages.

2.2 Principal components analysis

This volume describes the computation and usual graphical display of a normalised principal components analysis (PCA) processed on physical and chemical data (Carrel et al., 1986, Approche graphique de l'analyse en composantes principales normée : utilisation en hydrobiologie. *Acta Œcologica, Œcologia Generalis* : 7, 189-203) Alternative computation and graphical representations (canonical graphs, cartography of scores, data reconstitution) also are presented.

J.-M. Olivier. 36 pages.

2.3 Correspondence analysis

This volume describes the computation and usual graphical display of a correspondence analysis processed on bird count data (Auda, Y., & al., 1983, La dispersion spatiale des Oiseaux au cours du cycle annuel : deux méthodes de description graphique. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : III, 297, 387-392).

J.M. Olivier & S. Dolédec. 18 pages.

2.4 Multiple correspondence analysis

This is an introduction to the analysis of tables containing categorical (qualitative) data. In this case, values are represented by modalities. These modalities can be ordered resulting in an ordinal coding. In this volume we perform a multiple correspondence

analysis on a data set dealing with cat's fecundity (Pontier, 1984, Contribution à la biologie et à la génétique des populations de chats domestiques (Felis catus). Thèse de 3° cycle. Université Lyon 1. 1-145).

J.M. Olivier & S. Dolédec. 12 pages.

2.5 Fuzzy correspondence analysis

The paper deals with the multiple correspondence analysis extended to fuzzy coded disjunctive arrays. The example concerns aquatic Coleoptera species and was treated in the paper by Chevenet et al. (1994,A fuzzy coding approach for the analysis of long-term ecological data. Freshwater Biology : 31, 295-309).

S. Dolédec & F. Chevenet. 20 pages.

2.6 Between- and within-groups principal components analyses

In this volume, the statistical analysis of a multivariate environmental array is described. Quantitative variables collected at s locations for t sampling dates are analysed. To have a distinct view of the respective influence of the seasonal succession and the sample location on the variability of the measures, principal components analyses were used on tables from the linear model of variance analysis in a two-way layout with one observation per cell. Moreover, this volume introduces to the use of multivariate techniques using projection onto a subspace.

S. Dolédec & D. Chessel. 27 pages.

2.7 Structures à une seule dimension

Il arrive souvent qu'une analyse multivariée conduise à l'interprétation d'un seul facteur. La fiche regroupe la description des outils graphiques qui sont utiles dans cette circonstance.

D. Chessel & J. Thioulouse. 30 pages.

2.8 Analyse Non Symétrique des Correspondances

La fiche décrit le principe et l'usage de l'analyse non symétrique des correspondances créée par Lauro, N. & D'Ambra, L. (1984, *in* : Data Analysis and Informatics III. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 433-446). Le test de Light, R.J. & Margolin, B.H. (1971, An analysis of variance for categorical data. *Journal of the American Statistical Association* : 66, 534-544) lui est associée dans les deux options NSCA du module COA. L'enregistrement des occurrences d'espèces à partir des collections, de la littérature ou des relevés exploratoires est particulièrement en accord avec cette approche. L'asymétrie des relations espèces-relevés vues par l'averaging est abordée.

D. Chessel & C. Gimaret. 30 pages.

2.9 Questions-Réponses : Codage de modalités

Cette fiche est un compte-rendu de consultation statistique pour M. Le Berre, J. Reynaud-Feurly et H. Zucchetta (IREM de Lyon). Il décrit une approche possible des résultats d'une enquête, à partir des renseignement fournis par les auteurs. Il permettra de traiter quelques principes théoriques à partir de matériel concret dans un stage de formation destiné à des professeurs de mathématiques (corrélation, analyse en composantes principales, analyse des correspondances multiples, images euclidiennes). Cette première partie est consacrée à un examen détaillé des diverses parties du questionnaire.

D. Chessel 48 pages.

2.A Questions-Réponses : Correspondances multiples

Cette fiche utilise l'enquête menée par M. Le Berre, J. Reynaud-Feurly et H. Zucchetta (IREM de Lyon) pour décrire l'usage des modules utilisant des variables

qualitatives. Cette deuxième partie illustre les manipulations associées à l'analyse des correspondances multiples d'un tableau faiblement structuré. On souligne les précautions à prendre dans la manipulation des cartes factorielles de modalités.

D. Chessel et J. Thioulouse. 24 pages.

2.B Entre ACP et ACM : l'analyse de Hill et Smith

Les données écologiques sont fréquemment structurées en groupes de relevés en particulier dans le temps et dans l'espace. Les données morphométriques peuvent être aussi structurées en groupe de descripteurs. Ces descripteurs peuvent être qualitatifs ou quantitatifs. On les réunit par type pour des critères techniques bien que biologiquement une classe signifiante de descripteurs puissent comprendre des variables de plusieurs types. A partir d'un problème difficile d'analyse de données génétiques, on décrit l'usage de l'analyse de Hill & Smith (1976, Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* : 25, 249-255) qui permet de mélanger variables qualitatives et quantitatives.

D. Chessel & J.M. Lascaux. 30 pages.

2.C Analyse des correspondances avec colonne de référence

Quand une table de contingence contient une colonne de poids très élevé, cette colonne peut servir de point de référence. La distribution associée à la colonne de référence définit le poids des lignes, l'origine dans l'ensemble des profils colonnes et la métrique du Khi2 dans cet espace. L'inertie est alors une somme de Khi2 d'ajustement. La fiche donne donc un exemple d'analyse des correspondances sur modèles de B. Escofier (Analyse factorielle en référence à un modèle. Applications à l'analyse d'un tableau d'échanges. *Revue de Statistique Appliquée* : 32, 4, 25-36, 1984).

D. Chessel . 8 pages.

3 — Variables instrumentales

On trouve ici les méthodes classiques de régression (Régression linéaire multiple, sur composantes, locales) qui peuvent surtout inviter à l'utilisation de logiciels plus performants dans ce créneau. L'analyse en composantes principales sur Variables Instrumentales (ACPVI) et l'Analyse Canonique des Correspondances (ACC ou AFCVI) sont traîtées de manière apprentée aux objectifs du logiciel CANOCO [1]. Les méthodes de régression PLS de première et seconde génération sont disponibles.

3.1 Variables explicatives indépendantes

La fiche décrit la régression linéaire dans le cas le plus simple et le plus clair, celui de variables explicatives indépendantes. La première situation abordée est celle d'une variable explicative unique (Module UniVar). On considère ensuite k explicatives indépendantes (Module OrthoVar). La régression sur composantes principales (PCR) et la régression sur vecteurs propres de voisinages (NER) sont de ce type. On souligne que le principe de régression linéaire permet d'aborder des structures qui ne le sont pas du tout comme des courbes de réponses en cloche ou des chroniques auto corrélées négativement.

D. Chessel & J. Thioulouse. 28 pages.

3.2 Régression linéaire

La fiche décrit deux méthodes de régression linéaire (module LinearReg) dans le cadre d'un problème de prédiction d'une variable biologique par des variables d'environnement, problème posé dans un article récent de P. Baran & Coll. (1993 *Bull. Fr. Pêche Piscic* : 331, 321-340). On accorde une certaine importance à l'examen des variables initiales et à la définition de l'objectif visé. On aborde la régression multiple classique, les difficultés qu'elle soulève, et la solution proposée par la régression PLS, ou régression partiellement aux moindres carrés. Inventée en chimiométrie, dont elle est un standard méthodologique, la régression PLS gagne à être connue en écologie. L'algorithme utilisé est décrit par Ter Braak & Juggins (1993, *Hydrobiologia* : 269/270: 485-502, p. 487).

D. Chessel & J. Thioulouse. 26 pages.

3.3 Régression PLS de seconde génération

La fiche propose une introduction à la logique de la régression partiellement aux moindres carrés de deuxième génération (module PLSgen2). Quelques indications générales sont données sur des exemples reproductibles. Une comparaison des sorties du module PLSgen2 avec celles du logiciel SIMCA (http://www.umetri.se/simca-pg.htm) utilisé dans un ouvrage à paraître de M. Tenenhaus est détaillée.

D. Chessel & L. Monimeau. 26 p.

3.4 Analyses Discriminantes

Parmi de nombreuses méthodes de discrimination, l'analyse discriminante linéaire, ou analyse factorielle discriminante est la plus connue. Elle sert essentiellement à décrire ce qui distingue les moyennes par groupe de plusieurs variables mesurées sur plusieurs individus de plusieurs groupes. Si on ne poursuit pas cet objectif, la méthode est sans signification. La fiche décrit l'emploi du module Discrimin. En illustrant un problème de morphométrie avec l'analyse discriminante on remonte aux sources (Mahalanobis, P.C. (1936) On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* : 12, 49-55). Le programme est cependant beaucoup plus général et on l'emploie après une ACP à centrage multiplicatif, ce qui est inhabituel.

S. Dolédec, D. Chessel et H. Persat. 26 p.

3.5 Ordination sous contraintes

L'Analyse en Composantes Explicatives (ACE) d'Obadia (1978 - *Revue de Statistique Appliquée* : 24, 5-28) est la version ACP/ACP d'une méthode très générale de couplages de tableaux qu'on peut mettre en œuvre sur tout type de variables dans le module Projectors. La version AFC/ACP est l'Analyse Canonique des Correspondances ou AFCVI ou encore AFC sous contraintes. Diverses illustrations introduisent à la plasticité de ces méthodes et en soulignent les limites. Le module Projectors permet ainsi des analyses sous contraintes variées, en particulier les analyses canoniques des correspondances partielles, les analyses des covariances partielles, LONGI. Toutes les variantes sont exécutables sur variables qualitatives, quantitatives et floues.

D. Chessel. 42 pages.

3.6 Analyse Discriminante des Correspondances

L'Analyse discriminante des correspondances est une extension de l'analyse discriminante aux tableaux d'analyse des correspondances qui possède une propriété de double averaging optimal. On l'utilise pour des tableaux faunistiques à deux niveaux de relevés. On pourrait l'employer pour des tableaux faunistiques à deux niveaux de détermination.

D. Chessel & J. Thioulouse. 20 pages.

3.7 Analyse Canonique des Correspondances

L'Analyse Canonique des Correspondances de ter Braak (1986 - Canonical correspondence analysis : a new eigenvector technique for multivariate direct gradient analysis. *Ecology* : 69, 69-77) est la version ACP/AFC d'une méthode très générale de couplages de tableaux qu'on peut mettre en œuvre sur tout type de variables dans le module Projectors. Les illustrations portent sur des relevés avifaunistiques couplés à des relevés d'architecture de la végétation.

D. Chessel, S. Benyacoub & S. Dolédec. 38 pages.

3.8 Inertie intra-classes et mesures de diversité

Dans 25 stations on a mesuré 11 paramètres environnementaux en 10 points aléatoires. La variabilité stationelle définit la diversité mésologique. Dans ces 25 stations on a capturé 1008 poissons de 23 espèces. Ceci permet de définir la diversité taxonomique. Ces poissons portent 3 mesures morphométriques qui définissent la diversité stationelle de la taille des poissons d'une part et de la forme des poissons d'autre part. Une méthode générale de calcul de ces différentes composantes de la diversité est proposée (Discrimin : Within Parameters). Les relations entre ces composantes est analysées. La diversité de la forme est seule reliée à la diversité mésologique de manière complexe. L'indice de diversité proposé est valable pour tous les types de variable.

L. Hill, B. Statzner, J.M. Olivier & D. Chessel, 20 pages.

4 — Couplages et contraintes

Cette partie regroupe des méthodes récentes de couplages de deux tableaux alternatives au précédentes qui l'emporte en général en stabilité numérique. La documentation thématique est explicite :

4.1 Co-structure de deux analyses en composantes principales

La fiche aborde un des aspects du couplage de tableaux, celui du test de l'existence d'une co-structure entre deux tableaux individus-variables dont on a étudié la structure par une analyse en composantes principales. ADE propose un test de permutation sur une analyse de co-inertie basée sur l'analyse inter-batteries de Tucker (1958, An inter-battery method of factor analysis. Psychometrika : 23, 2, 111-136) connue aussi comme le premier pas des *Partial Least Square regression* (Höskuldsson, 1988, PLS regression methods. Journal of Chemometrics : 2, 211-228).

S. Dolédec & D. Chessel. 20 pages.

4.2 Profils écologiques et analyse de co-inertie

La fiche illustre les calculs de profils écologiques et l'exacte similitude entre AFC des profils écologiques (Romane, 1972, Un exemple d'utilisation de l'analyse factorielle des correspondances en écologie végétale. In : Grundfagen und Methoden in der Pflanzensoziologie. Maarel, E. van der & Tüxen, R. (Eds.) Dr. W. Junk b.v., The Hague. pp. 155-162.) et analyse de co-inertie d'un tableau de variables environnementales qualitatives et d'un tableau de végétation en présence-absence. Cette notion est étendue à des données de végétation quantitatives. La co-structure entre une AFC et une ACM est alors la version complète de l'AFC des tableaux de profils écologiques. L'illustration proposée est celle de Mercier et al. (1992, Complete

correspondence analysis of an ecological profile data table: a central ordination method. Acta Œcologica : 13, 25-44).

S. Dolédec, D. Chessel & P. Mercier. 28 pages.

4.3 Couplage d'analyses des correspondances sur codages flous

La fiche "Fuzzy correspondence analysis" décrivait les programmes abordant les tableaux de distributions de notes associant individus et variables qualitatives. On fait ici l'analyse de co-inertie de deux tableaux de ce type. L'exemple est tiré de Chevenet et al. (1994, A fuzzy coding approach for the analysis of long-term ecological data. Freshwater Biology : 31, 295-309).

S. Dolédec, D. Chessel & F. Chevenet. 20 pages.

4.4 Triplet d'analyses des correspondances

La fiche décrit une procédure d'analyse à trois tableaux utilisant un tableau d'AFC comme lien entre deux tableaux quelconques. Cette analyse est appelée analyse RLQ (R pour tableau en mode R, Q pour tableau en mode Q, et L tableau du lien) est basée sur les propriétés d'optimisation de la corrélation lignes-colonnes en AFC. Dans l'exemple traité les tableaux R et Q sont des tableaux de variables qualitatives et constituent donc un cas particulier (Dolédec, S., Chessel, D., Ter Braak, C.J.F. & Champely, S. (1996) Matching species traits to environmental variables: a new three-table ordination method. Environmental and Ecological Statistics : 3, 143-166).

D. Chessel, S. Dolédec & S. Champely. 64 pages.

4.5 Comparaison de deux tableaux homogènes de tests de toxicité

La fiche porte sur la description des tableaux homogènes quantitatifs, en particulier de leur centrage. Le centrage multiplicatif est le plus adapté pour des tableaux homogènes de tests de toxicité. On illustre une analyse de co-inertie entre deux ACP avec un centrage multiplicatif commun aux deux tableaux. On met en évidence, au delà de la similitude apparente, une différence de fond entre résultats de tests *in vivo* et *in vitro*, différence systématique exprimée par le biais d'une analyse de co-structure.

D. Chessel & J. Devillers. 22 pages.

4.6 Avant-après, amont-aval : les couples de tableaux totalement appariés

La fiche décrit les méthodes d'analyse des couples de tableaux complètement appariés (mêmes individus, mêmes variables). L'analyse de co-inertie prend dans cette situation des propriétés très particulière (Torre, F. & Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. Revue de Statistique Appliquée : 43, 109-121). Elle est complétée par une analyse des différences enrichie de nouvelles aides à l'interprétation. Des couples de tableaux de mesures acquises sur deux stations (Carrel, G. & Berget, J.C. , 1995. Suivi hydrobiologique du Rhône. Site de Cruas-Meysse (1982-1995). Cemagref Aix-en-Provence, Rapport à EDF, 1-44 + annexes.) ou à deux dates (Vespini, F., Légier, P. & Champeau, A. (1987) Ecologie d'une rivière non aménagée des Alpes du Sud : Le Buëch (France) I Evolution longitudinale des descripteurs physiques et chimiques. *Annales de Limnologie* : 23, 151-164) illustrent ces propositions.

D. Chessel & G. Carrel. 19 pages.

4.7 Niches écologiques et couplages de tableaux

La fiche décrit une méthode d'ordination et deux méthodes de couplage basées sur la notion de marginalité des niches écologiques (Hausser, J., 1995, *Säugetiere der Schweiz. Mammifères de la Suisse. Mammiferi della Svizzera*. Birkhäuser Verlag, Basel, Boston, Berlin. 501 p.). Voisines de l'analyse des correspondances et de l'Analyse Canonique des correspondances, ces procédures s'en éloignent fortement en cas de facteurs écologiques limitant (pollution, salinité) ou favorisant (météorologie)

l'abondance d'un groupe important de taxons. Le module Niche permet d'exécuter ces propositions.

D. Chessel & C. Gimaret. 33 pages.

4.8 Individus et informations supplémentaires

La fiche décrit l'usage des individus supplémentaires dans des circonstances variées. En particulier, cette pratique est étendue aux analyses inter et intra-classes et aux analyses de co-inertie. Une illustration porte sur l'analyse de co-inertie de deux analyses inter-classes. On distingue ce qui relève directement de la projection euclidienne de vecteurs quelconques sur des plans principaux de ce qui touche à l'illustration des plans factoriels par de l'information externe aux tableaux de données.

D. Chessel. 24 pages.

4.9 Analyses canoniques et listes d'occurrences d'espèces

La fiche précise la perturbation théorique introduite par le traitement des listes d'occurrences dans l'ensemble des méthodes linéaires d'analyse des données écologiques. Les méthodes d'analyses canoniques comme l'AFC, l'analyse canonique des correspondances et l'analyse canonique classique jouent un rôle central dans ce domaine. Les liens entre ces techniques et les méthodes dites à variables instrumentales sont abordés. Une révision globale de la place de l'analyse canonique en écologie semble nécessaire.

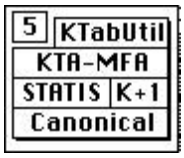
D. Chessel & C. Gimaret. 37 pages.

4.A Quand l'analyse canonique des correspondances est-elle une analyse canonique ?

La fiche montre comment l'analyse canonique des correspondances est une analyse canonique quand on confronte une liste d'occurrences d'espèces avec un tableau de variables environnementales portant sur ces occurrences. Cela modifie sensiblement les conditions d'utilisation de la méthode quand elle est employée, comme c'est le cas général, sur des tableaux relevés-espèces.

D. Chessel & C. Gimaret. 31 pages.

5 — Méthodes à *K* tableaux



ADE-4 propose ici un module complet d'analyse factorielle multiple [2] et une version généralisée de la méthode STATIS [3]. Un effort important a été fait pour mettre à disposition ces méthodes peu utilisées :

5.1 STATIS et valeurs typologiques des groupes faunistiques

La fiche donne une première présentation de la méthode ACT-STATIS (Lavit, Ch., 1988, *Analyse conjointe de tableaux quantitatifs*. Masson, Paris. 1-240). On aborde le cas d'un multi-tableau formé d'un tableau faunistique et de la répartition des taxons en plusieurs groupes faunistiques. Une mesure de la valeur typologique de chaque groupe est proposée.

D. Chessel, J. Thioulouse et M. Hanafi. 30 pages.

5.2 STATIS et ordination de plusieurs ensembles de relevés

Sur proposition de leur auteur, les données ichtyologiques publiées dans la thèse de J. Verneaux (1973 - Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon. 257 p.) sont revisitées avec la méthode STATIS. La comparaison de trois groupes de variables (géomorphologiques, ichtyologiques et physico-chimiques) permet de souligner l'objectif particulier qu'on se donne en utilisant cette méthode multi-tableaux.

D. Chessel, J. Thioulouse et M. Hanafi. 34 pages.

5.3 L'analyse factorielle multiple

L'Analyse factorielle multiple (AFMULT) est une des méthodes les plus pratiques parmi les méthodes d'ordination simultanée de *K* tableaux. La fiche reprend les principaux travaux des auteurs de l'AFMULT, B. Escofier et J. Pages, pour décrire l'utilisation du module MFA (option Variable groups) qui contient cette méthode.

D. Chessel. 27 pages.

5.4 L'analyse triadique partielle

La fiche décrit le mode d'emploi de l'analyse triadique partielle. La méthode, qu'on peut appeler STATIS sur les tableaux, par opposition à STATIS sur les opérateurs, permet de faire une analyse moyenne de plusieurs ACP portant sur les mêmes individus et les mêmes variables. Pour plusieurs AFC, dans les mêmes conditions, le problème des pondérations est résolu par l'AFC moyenne de Foucart (1978, Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des données* : 2, 67-84). Les données de Blondel et Farré (1988, The convergent trajectories of bird communities along ecological successions in european forests. *Œcologia (Berlin)* : 75, 83-93.) illustrent avec précision les questions qu'on peut ainsi aborder dans un cube de données.

D. Chessel, J. Thioulouse & M. Simier. 27 pages.

5.5 Analyses de co-inertie multiple : introduction

L'Analyse de co-inertie multiple (ACOM, Analyses de la co-inertie de K nuages de points. Revue de Statistique Appliquée : 44, 35-60) généralise à K tableaux l'analyse de co-inertie (Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294) pour deux tableaux de la même manière que l'analyse canonique généralisée de Carrol (A generalization of canonical analysis to three or more sets of variables. *Proceeding of the 76th Convention of the American Psychological Association* : 3, 227-228) généralise l'analyse canonique classique de Hotteling (Relations between two sets of variates. *Biometrika* : 28, 321-377). La fiche décrit l'utilisation du module KTA qui contient cette méthode.

D. Chessel & M. Hanafi. 24 pages.

5.6 ACOM et ordination de *K* ensembles de relevés

On compare l'analyse de co-inertie multiple (ACOM) à la méthode STATIS en testant sa capacité à ordonner simultanément *K* ensembles de relevés. On utilise pour ce faire les données hydrobiologiques de la thèse de J. Verneaux (Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon. 1-257). Les trois exemples apportent une compréhension plus fine du rôle des deux méthodes. L'ACOM fait K analyses (coordonnées) et en trouve une moyenne. STATIS cherche une moyenne et en fait une analyse. Si chaque analyse est pertinente et mérite un examen approfondi, l'ACOM se justifie. Si chaque analyse est d'intérêt limité STATIS est plus clair. Dans la comparaison de peu d'études régionales comportant beaucoup de relevés, on pourra préférer l'ACOM. Dans la comparaison de nombreuses études régionales comportant un nombre limité de relevés, on devra s'en tenir à STATIS. Les deux méthodes sont

centrées sur la notion de structures communes et les résultats communs qu'elles donnent avec des logiques différentes sont normalement cohérents.

D. Chessel & M. Hanafi. 21 pages.

5.7 STATIS-AFMULT-ACOM Cas des mêmes individus

La fiche résume la mise en œuvre des trois méthodes STATIS, AFMULT et ACOM. *K* tableaux portent sur les mêmes individus. On définit la structure des données et le centrage (**KTabUtil**). Le listing de sortie (**STATIS-MFA-KTA**)est reproduit et la signification du contenu des fichiers est brièvement rappelée. Les principales figures associées aux critères optimisés sont construites. On peut utiliser cette fiche comme mémento pour une mise en œuvre rapide. Les illustrations portent sur les données de L.E. Friday (1987, The diversity of macroinvertebrate and macrophyte communities in ponds. Freshwater Biology : 18, 87-104).

D. Chessel & M. Hanafi. 26 pages.

5.8 STATICO : STATIS et Co-inertie

La fiche décrit le mode d'emploi de l'analyse triadique partielle étendue aux opérateurs de co-inertie. La méthode, qu'on peut appeler STATICO, est une analyse triadique partielle sur les tableaux croisés dans les analyses de co-inertie. On combine alors la logique de STATIS (trouver ce qui dans plusieurs tableaux constitue le fond typologique commun, fond commun qui peut exister derrière de fortes particularités propres à chaque tableau) et la logique des couplages de co-inertie (trouver ce qui dans deux groupes de descripteurs engendre une typologie commune des objets décrits dans chacun des tableaux). Cette approche fait suite à une demande de l'équipe de D. Nandris (Phytopathologie, ORSTOM, Nouméa).

D. Chessel, M.Simier, M. Hanafi. 41 pages.

5.9 Analyses des correspondances et *K*-tableaux

La fiche propose une introduction de la logique de l'analyse des correspondances dans l'usage des analyses à *K*-tableaux (STATIS et AFMULT). On présente un jeu de données (6 campagnes de chalutage dans le golfe du Lion) qui demande une approche analyses des correspondances des structures spatio-faunistiques. La simple utilisation de l'option CoaKTab dans KTabUtil permet l'usage de KTA, STATIS et AFMULT dans cette logique AFC. Des logiques internes très différentes conduisent à des résultats très proches.

D. Chessel & J.C. Gaertner. 24 pages.

5.A Instabilité de descripteurs et stabilité de structures en végétation aquatique

La fiche réunit un ensemble de pratiques de description d'un cube de données floristiques. Une première partie présente les données traîtées. La seconde effleure le nombre considérable d'approches possibles et compare les trois analyses simples de base. La troisième compare les différentes approches possibles du compromis temporel. La notion de structure spatiale stable à composantes instables est résumée par une analyse de co-inertie multiple.

D. Chessel & G. Bornette. 38 pages.

5.B Analyse d'un tableau de taux

On est souvent confronté en épidémiologie à l'analyse de tableaux dont les termes sont des taux. Lorsque les dénominateurs sont inégaux d'une ligne à l'autre, les centrages "colonne" classiques de l'ACP centrée ou normée ne sont pas autorisés car la moyenne d'une série de taux n'est pas une moyenne arithmétique. Pour l'analyse de base d'un tel tableau, on donne ici un exemple d'utilisation de l'analyse factorielle des correspondances décentrée (AFCD), extension de l'AFC simple, dans laquelle la marge "ligne" est remplacée par une marge extérieure imposée, contenant les effectif des dénominateurs. La structure spatio-temporelle du tableau peut ensuite être décrite par

l'utilisation de méthodes déjà proposées dans une logique d'AFC (analyses interclasse, intraclasse, sur variables instrumentales, STATIS, analyse de co-inertie multiple). On montre que la structure de la programmathèque ADE-4 permet d'étendre les analyses K-tableaux à toutes les options d'analyse de base.

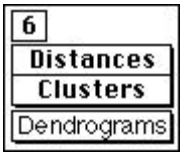
E. Tillard & B. Faye. 50 pages.

5.C ACOM et mesure de la valeur typologique

L'Analyse de co-inertie multiple (ACOM) généralise à K tableaux l'analyse de coinertie pour deux tableaux. On la compare ici à STATIS en testant sa capacité à mesurer la valeur typologique des groupes faunistiques. Des logiques d'analyse distinctes conduisent à des résultats cohérents.

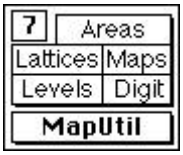
M. Hanafi & D. Chessel. 14 pages.

6 — Distances et partition



Cette partie ne possède pas de documentation thématique, les fonctions assurées étant décrites dans les ouvrages d'analyse des données [4]. Les fiches des modules sont explicites des services que peuvent rendre ces outils. Le module distances permet de calculer 24 types de matrices de distance utiles en classification ou en analyse en coordonnées principales. Le module Clusters calcule cinq types classiques de hiérarchie et propose la classification "autour des centres mobiles".

7 — Cartographie



Cette partie n'est disponible que sur MacIntosh. Les fiches des modules décrivent les fonctions que peuvent rendre ces modules qui peuvent servir à des illustrations sans prétendre avoir des fonction de SIG. Est ajoutée à cette partie de la documentation les fonctions du module NGStat, les analyses multivariées sous contraintes spatiales utilisant largement les outils cartographiques.

7.1 Enregistrements de l'information spatiale

La fiche décrit les objets principaux et les fonctions associés à l'enregistrement des données spatiales. On a ainsi une vue d'ensemble des modules Areas, Digit, Lattices, Levels, Maps et MapUtil.

D. Chessel. 32 pages.

7.2 Analyses des données de radiopistage (travaux de L. Blanc)

Le radiopistage est devenu une technique fréquemment utilisée en écologie pour approcher l'occupation de l'espace par un individu. Sont traités les principaux objectifs visés au cours des études hydrobiologiques : détermination de la localisation d'un animal, étude des mouvements à différentes échelles spatio-temporelles, estimation de

la taille du domaine vital et analyse de l'habitat à l'intérieur de ce domaine vital. Les méthodes d'étude ainsi que les problèmes statistiques associés à ces méthodes, sont exposés et commentés. Un module intitulé "Radiotracking" a été écrit pour une première analyse des données brutes recueillies par radiopistage et pour représenter sur un fond de carte les localisations estimées. Ce programme crée également des fichiers avec lesquels on accède directement à deux autres logiciels téléchargeables sur le réseau : ANTELOPE et WILDTRAK Ces derniers nous ont permis d'obtenir par différentes méthodes des estimations des aires d'activités journalières des poissons étudiés par L'ORSTOM dont nous avons superposé les représentations sur un fond de carte.

L. Blanc. 84 pages.

7.3 Autocorrélation et composantes cartographiables

La fiche donne quelques illustrations des notions élémentaires de variance locale et d'autocorrélation spatiale à la base de l'analyse des structures spatiales. Ces deux notions utilisent des graphes de voisinage. Le test de Geary, l'AFC des matrices de graphe, la diagonalisation des opérateurs de Moran et l'analyse multivariée basée sur la maximisation de l'autocorrélation sont des éléments préparatoires aux méthodes multimétriques.

D. Chessel, J. Thioulouse et S. Champely. 26 pages.

7.4 Analyse d'inertie locale

La fiche décrit l'usage des analyses d'inertie locale dans plusieurs circonstances. D'un principe relativement simple, cette analyse peut donner des résultats semblables à ceux obtenus d'une analyse simple dans le cas de gradients multivariés. Dans d'autre cas on obtiendra des résultats complètement différents. Ce type d'analyse concentre l'attention sur la variabilité entre individus voisins et se place comme radicalement antinomique à la recherche des composantes cartographiables d'une enquête écologique à relevés régionalisés. Les exemples traités montrent comment structure locale et globale peuvent se superposer ou se mélanger dans une analyse simple.

D. Chessel et J. Thioulouseet & S. Dolédec. 26 pages.

On insiste sur les interfaces avec d'autres logiciels en freeware sur Internet comme dans NGUtil avec :

dataThief 2.0 de Kees Huyser et Jan van der Loan [5]

dataThief 2.0 de Kees Huyser et Jan van der Loan [6]
DigiGraf 1.4.4 de Simon Tortike [7]

dans MapUtil avec :

Cabri-Graphe de l'INRIA [8]
Groups & Graphs de Bill Kocay [9]

et dans RadioTracking avec :

ANTELOPE 1.3 de J. Bradbury et S. Vehrencamp [10]
WILDTRAK de Ian Todd [11].

Références

[1] Ter Braak, C.J.F. (1987c) CANOCO - a FORTRAN program for Canonical commnity ordination by [partial][detrended][canonical] correspondence analysis and redundancy analysis. Software documentation. Version 2.1, TNO Institute of Applied Computer Science, Wageningen. http://vax.ox.ac.uk/HELP/@OX$USERLIB:[HELP]USERLIB/CANOCO31/DISTRIBUTION

[2] Escofier, B. & Pagès, J. (1994) Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis* : 18, 121-140.

[3] Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis* : 18, 97-119.

[4] Lebart, L., Morineau, A. & Piron, M. (1995) Statistique exploratoire multidimensionnelle. Dunod, Paris. 1-439.

Manly, B.F. (1994a) Multivariate Statistical Methods. A primer. Second edition. Chapman & Hall, London. 1-215.

Digby, P. G. N. & Kempton, R. A. . (1987) Multivariate Analysis of Ecological Communities. Chapman and Hall, Population and Community Biology Series, London. 1-205.

[5] http://www.nikhefk.nikhef.nl/~keeshu/dataThief.sit.hqx

[6] http://www.nikhefk.nikhef.nl/~keeshu/dataThief.sit.hqx

[7] mmpe.mineral.ualberta.ca dans le répertoire ftp/pub/digigraf144.hqx

[8] http://www-cabri.imag.fr/CabriGraphes/cabri_graph.html

[9] http://130.179.24.217/G&G/G&G.html

[10] http://www-biology.ucsd.edu/research/vehrenbury/programs.html

[11] http://www.uni-sb.de/philfak/fb6/fr66/tpw/telem/software/wildtrak.htm